

Investigating prediction in L2 morpho-syntax: A visual world study

by

Teljer Liburd

B.A., Emory University, 2007

M.A., Georgia State University, 2011

Submitted to the Graduate Faculty of the
Kenneth P. Dietrich School of Arts and Sciences in fulfillment
of the requirements for the degree of
Master of Science in Cognitive Psychology

University of Pittsburgh

2014

UNIVERSITY OF PITTSBURGH

Dietrich School of Arts and Sciences

This thesis was presented

by

Teljer Liburd

It was defended on

August 8, 2014

and approved by

Tessa Warren, Associate Professor, Psychology Department

Natasha Tokowicz, Associate Professor, Psychology Department

Michael Dickey, Assistant Professor, Communication Sciences and Disorders Department

Thesis co-Directors: Tessa Warren and Natasha Tokowicz,

Associate Professors, Psychology Department

Copyright © by Teljer Liburd

2014

Investigating prediction in L2 morpho-syntax: A visual world study

Teljer Liburd, M.A, M.S.

University of Pittsburgh, 2014

When speakers of gendered languages hear determiners, they anticipate nouns that share the determiner's gender. We examined whether beginning L2 learners anticipate upcoming nouns using determiners' number/grammatical gender, as a function of 1) cross-language similarity and 2) reliability in mapping of the determiner to an upcoming noun. Native English speakers were taught Dutch nouns, and determiners that were "Similar" or "Different" in English and Dutch, and "Unique" to Dutch. Half the participants were taught determiners that had reliable, one-to-one mapping to upcoming nouns, and the other half was taught a determiner that mapped to more than one type of noun. We tracked eye movements to pairs of pictures while participants listened to partial determiner-final Dutch sentences; they identified by button press which picture best completed the sentence. Accuracy was higher for Similar than Different and Unique sentences, and reaction time (RT) was faster for Similar than Different and Unique sentences. Cross-language similarity also influenced how quickly participants looked to the target after hearing the determiner. Reliability effects were most evident in RT data, which showed that participants in the High reliability group responded more quickly than those in the Low reliability group. Cross-language similarity appears to modulate the learnability of mapping a determiner to its noun, suggesting that beginning L2 learners can use morpho-syntax to make predictions during online sentence comprehension.

TABLE OF CONTENTS

ACKNOWLEDGEMENTS	ix
1.0 INTRODUCTION.....	1
1.1 CURRENT STUDY	9
2.0 METHOD	11
2.1 Design	11
2.2 Cross-language Similarity	11
2.3 Reliability.....	13
2.4 Participants.....	14
2.5 Training Stimuli	14
2.6 Training Procedure.....	15
2.7 Eye-tracking Stimuli.....	16
2.8 Eye-tracking Procedure.....	11
3.0 RESULTS	19
3.1 BEHAVIORAL RESULTS.....	19
3.1.1 Accuracy	19
3.1.2 Reaction Time.....	20
3.2 EYE-TRACKING RESULTS.....	21
3.2.1 First-fixation Latency	21

3.2.2	Time Course	23
4.0	GENERAL DISCUSSION	26
	APPENDIX A	33
	APPENDIX B	34
	BIBLIOGRAPHY	35

LIST OF TABLES

Table 1. Cross-language similarity and reliability conditions.....	12
Table 2. Summary of study procedures	16
Table 3. Mean accuracy by Similarity and Reliability condition	19
Table 4. Mean reaction time (ms) for Similarity and Reliability condition.....	26
Table 5. List of word stimuli.....	33

LIST OF FIGURES

Figure 1. Visual representation of the High and Low reliability mappings of determiner “de”...	13
Figure 2. Mean accuracy and Similarity condition.....	20
Figure 3. Mean reaction time to elicit a button press after determiner offset by Similarity and Reliability condition.....	21
Figure 4. Mean first fixation latency to look at target from determiner offset for each Reliability and Similarity condition.....	22
Figure 5. Cumulative proportion of all trials on which participants looked to the target picture after determiner offset for each Similarity and Reliability condtion	23
Figure 6. Cumulative proportion of correct trials on which participants looked to the target picture after determiner offset for each Similarity and Reliability condtion	24

ACKNOWLEDGEMENTS

I would like to thank everyone that who supported me in the process of completing of my Master's. I give special thanks to my closest friends, especially Janelle, Yema, Miheret, and Angie who showered me with unwavering support and love throughout the entire process. Their encouragement was instrumental in me completing my thesis.

I would like to acknowledge Sigma Xi and Dietrich School of Arts and Sciences' Diversity Summer Research Fellowship who provided the funding for this research project.

Thank you to Anna Vazeou-Nieuwenhuis, the only Dutch native speaker I knew in Pittsburgh who graciously recorded all the stimuli, as well as thanks to Meghan Bathgate who recorded the English stimuli.

And of course, I am extremely grateful to my co-advisors Drs. Tessa Warren and Natasha Tokowicz for their academic rigor and scientific genius in guiding me in the development, execution, and completion of this thesis. With their support, I had the tools I needed to be successful. I would also like to extend a special thanks to Dr. Michael Dickey, a member of my Master's committee for all his thought-provoking insight and comments on the project, as well as Dr. Charles Perfetti for serving as chair of my mentoring committee and for all his advice and support.

1.0 INTRODUCTION

There is growing evidence that first language (L1) comprehension involves making predictions of likely upcoming material at multiple levels of analysis using, for example, morpho-syntax (e.g., Lew-Williams & Fernald, 2007), world knowledge (e.g., Borovsky, Elman & Fernald, 2012; Kamide, Altmann & Haywood, 2003), phonology (e.g., Brunellière & Soto-Faraco, 2013), and context (e.g., Van Berkum et al., 2005). These studies suggest that in L1, comprehenders draw on different types of information during online language comprehension, and then use that information to project expectations for what might come next. However, the evidence from previous research as to whether second language (L2) learners predict upcoming material online is inconclusive. The ability of L2 learners to use linguistic information to make predictions can provide an index of learning. This is because in order to use an L2 structure to predict upcoming material, a learner needs to have learned how that structure is implemented in the L2, know potential upcoming material that can follow that structure, and know something about the nature of the mapping between that structure and the upcoming material.

The current study investigates whether L2 learners make predictions of upcoming information on-line using morpho-syntax (number and grammatical gender). Specifically, we examine whether any ability to make predictions might be modulated by: 1) similarities and differences in how linguistic structures are instantiated across languages, and 2) how reliably these morpho-syntactic structures map to upcoming information. Knowing whether L2 learners

use morphosyntactic information to make predictions will bring insight into on-line L2 sentence processing and L2 morpho-syntactic learning.

Previous L1 research shows that linguistic information from verbs can immediately drive predictions of upcoming arguments. For example, Altmann and Kamide (1999) demonstrated that upon hearing a verb like *eat*, listeners anticipate that an edible argument will follow, even before hearing it. Participants looked at scene images—such as one with a boy sitting with some toys (a train, a truck and, a ball) and a cake nearby—while listening to sentences like: “The boy will eat the ____” or “The boy will move the ____.” The verb “eat” imposes a semantic constraint on upcoming arguments, meaning that only an edible argument can follow it (balls and trains are not edible), whereas *move* does not impose these restrictions (each of these objects is movable). They found that participants launched eye movements to the target (i.e., the cake) sooner in the “eat” condition than in the “move” condition. Moreover, in the “eat” condition eye movements were launched *before* the onset of “cake”, whereas in the “move” condition eye movements were launched *after* the onset of “cake.” These findings suggest that linguistic information, when informative, can contribute to making predictions.

Studies show that for native speakers of gendered languages (e.g., Spanish or French), gender information provides strong cues to upcoming nouns (Lew-Williams & Fernald, 2007; Valdés Kroff et al., 2008). Lew-Williams and Fernald (2007) showed that upon hearing a determiner, even child native Spanish speakers anticipate nouns that share that determiner’s gender. They had monolingual Spanish-speaking two- to three-year olds look at pairs of pictures while they listened to short sentences referring to one of two objects (e.g., *Encuentra la/el* ____ “Find the ____”). The pairs of pictures were either of the same gender (e.g., *la pelota* and *la galleta*) or of different genders (e.g., *la pelota* and *el zapato*). In Spanish, a determiner’s gender

and number is determined by the noun that follows it. Eye movement results showed that after the onset of the determiner the children oriented to the target picture faster on different-gender trials than on same-gender trials. Adult controls also evidenced facilitated prediction on the different-gender trials and were even faster to identify the target. This pattern of looking suggests that the gender information on the determiner was an informative cue and was used to help participants anticipate which noun was to follow. This is evidence that online comprehension in gendered languages involves rapid use of morpo-syntactic information.

Event-related potentials (ERP) have previously been used to examine implicit processing of gender information. One such study by Van Berkum et al. (2005) suggests that native speakers of gendered languages also use context to anticipate nouns of a particular gender. The authors presented native Dutch speakers with mini-stories that provided context for predicting upcoming nouns. Participants heard sentences with indefinite noun phrases that included an adjective (e.g., *een groot_{neuter} schilderij_{neuter}*, “a big painting”). In Dutch, indefinite determiners are not gender-marked. However, adjectives (which are prenominal) are marked for gender (common or neuter) based on the noun that modifies them (e.g., *groot_{neuter}/grote_{common}*, “big”). They found that, as early as on the adjective, participants elicited a small but reliable N400 when the gender of that adjective was inconsistent with gender of the upcoming noun they expected to hear, based on context. The N400 is a time-locked negative going ERP waveform that peaks around 400ms after stimulus onset, and is thought to reflect semantic processing more generally, but has been found to be an implicit response to semantically anomalous words. Strikingly, this deflection in the ERP waveform was detected as early as on the vowel of the adjective, which either had long (*groot* [ɣro:t]) or short (*grote* [ɣrɔ:tə]) durations. These results indicate that predictions can be made or disconfirmed at the earliest point at which disambiguating information is available. In

this case, even differences in the vowel durations were sufficient to confirm or disconfirm whether the upcoming noun was consistent with the prediction.

There is also evidence that gender information on determiners facilitate lexical access of upcoming nouns (Bates et al., 1996; Dahan et al., 2000; Grosjean et al., 1994; Wicha et al., 2004). In an eye tracking study by Dahan et al. (2000), native French speakers looked at an array of four pictures while listening to sentences asking them to click on one of them (e.g., “Click on the ____”). They found that when the determiner provided gender information (e.g., *le_{masculine} bouton_{masculine}*), participants had more fixations on the target and fewer fixations on cohort competitors compared to when the determiner was gender-ambiguous (e.g., *les_{neutral} bouteilles_{masculine}*). Bates et al. (1996) showed this same gender facilitation of lexical access with adjective-noun pairs. In particular, they found that Italian speakers were faster at auditory naming and grammaticality judgment when the adjective’s gender matched that of the noun compared to when it did not. These studies suggest that gender information on determiners and adjectives carry information that facilitate the access of upcoming nouns, and are therefore strong cues for making predictions.

It is clear that native speakers of gendered languages are sensitive to gender cues, which appear useful for making predictions. Some have, however, questioned whether L2 learners are sensitive to morpho-syntactic information on determiners, and whether they can use this information on-line to make predictions. Replicating their 2007 study, Lew-Williams and Fernald (2010) had intermediate proficiency adult L2 Spanish learners (self-rated proficiency of about 3.5 out of 5) listen to sentences referring to one of two pictures. They examined the time course of looking in same- versus different-gender trials using familiar Spanish nouns (Experiment 1) and novel Spanish nouns (i.e., Spanish non-words; Experiment 2). No training

was given for familiar nouns. In Experiment 2, four novel nouns were trained; two were made “masculine” and two “feminine” by pairing them with the definite determiners *el* and *la*, respectively, and by giving them *-o* and *-a* morphological endings (e.g., *la catela*, *el durino*). During training, participants saw a picture of a novel object while listening to sentences like *¡Mira, es la catela!* ‘Look, it’s the *catela*!’ Eye movement results for the L2 group showed no difference in looking for same- versus different-gender trials for the familiar nouns. However, for novel nouns, the L2 learners oriented to the target picture marginally more quickly on different-gender trials than on same-gender trials.

More recently, Grüter, Lew-Williams, and Fernald (2012) replicated this study with more advanced L2 learners (self-rated proficiency of 8.5 out of 10). The L2 group showed no difference in proportion of looking across same- versus different-gender trials for familiar nouns. However, both L1 natives and the L2 learners seemed to make use of the information on the determiner for the novel nouns – both groups showed a marginally significant difference in looking across trial type. One interesting thing to point out about this study is the time course in orienting to the target picture. Although both groups showed differences in looking across trial type, this difference was evident much later for the L2 group. This suggests that, compared to native speakers, L2 learners might be slower in their use of the determiner as a predictive cue. The results from these two studies bring insight into online processing in an L2 context and provide at least some evidence that L2 learners use the information on determiners as predictive cues.

The interesting finding in the studies by Lew-Williams and colleagues (2007, 2010) was that minimal explicit training on new determiner-noun pairs resulted in some evidence of prediction in L2 learners, whereas this was not the case for words they were already familiar

with. Based on differences in eye movement results for familiar versus novel nouns, it seems that strengthening participants' explicit knowledge of the determiner-noun pairings resulted in more efficient implicit use of information during comprehension. It appears that stronger explicit knowledge of these mappings, improved the automaticity of implicit eye movements in making a guess of what might come next. Thus, predictions in an L2 context can be considered a measure of L2 learning. That is, in order to take advantage of a linguistic determiner as a cue to an upcoming noun, a learner needs to have learned not only how that determiner is instantiated in the L2, but must also know what types of nouns can follow, as well as something about the nature of the relationship or mapping between the determiner and upcoming noun. It is this success in morpho-syntactic learning and how it contributes to driving predictive eye movements that interests us in this study.

If a learner's ability to make predictions says something about how well they learned to map one structure onto another, then it is important to consider factors that might make these mappings easier or harder to learn. Psycholinguistic models offer varying perspectives on how adults learn L2 morpho-syntax. One such model is the Unified Competition Model (UCM) (MacWhinney, 2005), which posits that learners rely on processing strategies from L1. This mechanism of cross-language transfer can help or hinder learning, depending on whether the L2 feature being processed is similar or dissimilar to the L1 feature driving the transfer. According to the UCM, linguistic structures in the L2 that work similarly in the L1 benefit from transfer from the L1, and are therefore easiest to learn. The model also describes structures that exist in both the L1 and L2, but are implemented differently in both languages; these types of features give rise to competition and are the most difficult to learn because of the mismatch in how they work across the two languages (e.g., Dutch has singular and plural markings on the definite

determiner, whereas English does not). Features unique to the L2 (i.e., that do not exist in the L1) give rise to neither transfer nor competition.

Another component that the UCM posits as important to learning a linguistic structure is its reliability as a cue. A linguistic structure is more or less reliable, depending on how consistently it cues upcoming material. For example, in English the determiner *this* always cues a singular upcoming noun, whereas the determiner *the* cues either singular or plural nouns. Therefore, *this* is a more consistent or more reliable cue to a singular noun than is *the*, and so *this* is a better cue to singularity than is *the*.

The UCM posits that cue reliability and, therefore, reliance on certain linguistic cues can differ across languages depending on how strong the cues are in their respective languages. For example, in English, word order is a dominant cue to identifying an actor in a sentence whereas in Dutch, case inflection is a strong cue for identifying an actor. Previous research has shown that although learners tend to initially rely on cues from the L1 when comprehending the L2, increased exposure to the L2 allows learners to shift their reliance to cues that are more suitable for use in the L2 (McDonald, 1987).

Several studies have addressed cross-language similarity and provide evidence that both transfer and reliability are important in L2 morphosyntactic learning (e.g., Foucart & Frenck-Mestre, 2011; Gillon Dowens et al., 2011; Tokowicz & MacWhinney, 2005; Tokowicz & Warren, 2010). However, these and similar studies have mostly tested learning based on sensitivity to morpho-syntactic violations. For instance, in an ERP study (Tokowicz & MacWhinney, 2005), native English speakers showed implicit sensitivity to grammatical violations in L2 Spanish, particularly for violations of grammatical constructions instantiated similarly across L1 and L2, but not for those instantiated differently. These results are consistent

with the predictions of the UCM in that they provide evidence of transfer based on morpho-syntactic similarity across languages. The present research extends the existing literature by using predictions to measure morpho-syntactic learning.

In our examination of predictions, it is relevant to consider the timing within which predictions might occur. Evidence from Grüter et al. (20012) suggests that L2 learners might not be as efficient as L1 speakers in processing information on gender-marked determiners. This might be expected given that L1 speakers' implicit processing is more automatized. It is therefore reasonable to hypothesize that L2 learners' underlying implicit processing might be slow compared to that of L1 speakers who have had a lifetime of exposure with determiner-noun mappings. If it is the case that L2 learners are just slower in their predictive eye movements, then a paradigm that can capture these slower predictions is needed.

A recent study by Mack et al. (2013) used a paradigm that captured slow predictions in people with aphasia. In their study they looked at integration versus prediction of an upcoming noun using verbs. While looking at a 4-picture array, people with aphasia listened to sentences with a restrictive (e.g., break) or nonrestrictive (e.g., open) verb with either an overt or covert noun (e.g., Susan will break/open the book vs. Susan will break/open the ____). They found that in sentences with an overt noun, aphasic participants were slower to orient to the target than age-matched controls, meaning that eye movements to a target occurred within a time window that coincided with the presentation of the noun. However, when the noun was covert, people with aphasia were able to make use of information on the verb to make a prediction of a target noun, but with a significant delay. Fixations to the target began *after* verb offset rather than during the processing of the verb, which is when the control group began fixating on the target. So although the presence of an overt noun aided in integration of a verb with the noun, removing

the final noun allowed people with aphasia more time to generate a prediction. Presuming that L2 learners will be slow in their prediction as well, we utilize a similar paradigm to that of Mack et al. (2013).

1.1 Current Study

The current study investigates how beginning L2 learners use morpho-syntactic information while comprehending. We address the following research question: Do L2 learners use information from determiners, particularly number and gender, to predict upcoming nouns, and if so, what is the time course of their predictions? To do this, we implemented a training study in which we trained participants on a small set of Dutch vocabulary, with particular focus on determiners and adjectives. Predictions were examined using eye tracking in the Visual World Paradigm. Using the framework of the UCM, the current study examines whether L2 predictions are modulated by 1) similarities or differences in how morpho-syntactic structures are instantiated across languages, and 2) the reliability in how morpho-syntactic structures map onto upcoming nouns.

The present research is important because whereas previous studies have addressed cross-language transfer and reliability separately, the current study examines these two mechanisms simultaneously. According to the UCM, both transfer and reliability should affect learning; considering these together allows the examination of their combined effect on learning. This will be key for structures unique to the L2. Because there is no transfer for structures unique to the L2, learning, should depend on the reliability of these structures as well their availability in the

input. A similar testing structure as Lew-Williams and Fernald (2010) was used except that explicit training on determiners was provided, as well as more extensive training on nouns. In doing so, we sought to increase the likelihood that participants learned the determiner-noun mappings, rather than just learning them as separate lexical items. A wider variety of determiners was used to capture the fact that nouns are typically paired with many different types of determiners, and also taking into account that determiners are instantiated differently across languages.

2.0 METHOD

2.1 Design.

This experiment used a 3 x 2 mixed design in which we manipulated cross-language similarity of determiner (Similar, Different, and Unique) as a within subject variable, and reliability of determiner-noun mappings (High versus Low) as a between subjects variable.

2.2 Cross-language Similarity.

Each participant was trained on determiners that are instantiated similarly and differently across Dutch and English, as well as on those that are unique to Dutch (see Table 1 for example sentences). Demonstrative determiners *dit* and *deze* ('this' and 'these') were trained and are considered *Similar* because of the likeness in how both languages use them to designate plurality. These were trained in the neuter gender only. The definite determiners *het* and *de* ('the_{singular}' and 'the_{plural}') are considered to be *Different* because although both languages use the definite article, Dutch has singular and plural markings on the definite article, whereas English does not; these will also be trained in the neuter only. *Similar* and *Different* determiners were identical in both the *High* and *Low* reliability conditions; the *Unique* determiners distinguished the two reliability conditions.

Table 1. *Cross-language similarity and reliability conditions.*

Similarity	High Reliability	Low Reliability
Similar	$\text{Dit}_{\text{neuterSING}} - \text{Deze}_{\text{neuterPLUR}}$ (Demonstrative determiners: This – These)	
	<i>Nicolaas ziet dit boek/deze boeken</i> ‘Nicholas sees this book/these books’	
Different	$\text{Het}_{\text{neuterSING}} - \text{De}_{\text{neuterPLUR}}$ (Definite determiners: The _{SING} – The _{PLUR})	
	<i>Nicolaas koopt het schilderij/de schilderijen</i> ‘Nicholas buys the painting/the paintings’	
Unique	$\text{Een rode}_{\text{commonSING}} / \text{Een rood}_{\text{neuterSING}} + \text{noun}$ (Indefinite noun phrase + adjective: A red + noun)	$\text{De}_{\text{commonSING}} - \text{Het}_{\text{neuterSING}}$ (Definite determiner: The _{SING} – The _{SING})
	<i>Nicolaas raakt een rode schildpad/een rood schilderij</i> ‘Nicholas touches a red turtle/a red painting’	<i>Nicolaas raakt de schildpad/het schilderij</i> ‘Nicholas touches the turtle/the painting’

Two different types of *Unique* determiners were trained, one for each reliability condition, both instantiating grammatical gender, which is present in Dutch but not in English. For Unique determiners in the High reliability condition, participants were trained on Dutch indefinite noun phrases consisting of the indefinite determiner *een* (‘a’) and the adjectives *rood* or *rode* (‘red’). The rule in Dutch is that when an indefinite noun phrase is singular and modified by an adjective, the adjective has to agree in gender with the noun. For singular noun phrases in the common gender the adjective *red* takes on an “e” inflection and becomes *rode*, and for singular noun phrases in the neuter gender, it remains as *rood*. This is unique to Dutch because, unlike in English, the adjective has to agree with the upcoming noun’s gender. For Unique determiners in the Low reliability condition, participants were taught the definite determiners *de* and *het* (the_{common} and the_{neuter}), which reflect the gender of their noun.

2.3 Reliability.

In the Low reliability condition, participants were taught two different mappings of the determiner *de*. That is, they were taught that *de* modifies a neuter plural noun (*Different*; reflects plurality), but they were also taught that *de* modifies a common singular noun (*Unique2*; reflects grammatical gender). Based on the predictions of the UCM, it was therefore expected that the low reliability in mappings of this determiner would make it more difficult to learn, compared to other determiners that only map onto one type of noun. Given that *de* was taught as a *Different* and *Unique* determiner in the Low reliability condition, its multiple mappings should affect learning in the Low reliability condition in both of these cross-similarity conditions. In contrast, in the High Reliability condition, participants were taught only one mapping for the determiner *de*—one that maps only to a neuter plural noun (*Different* determiner). Given its consistent mapping to the neuter plural, *de* as taught in High reliability is a more reliable cue to the neuter plural than *de* in the *Low* reliability condition (see Figure 1).

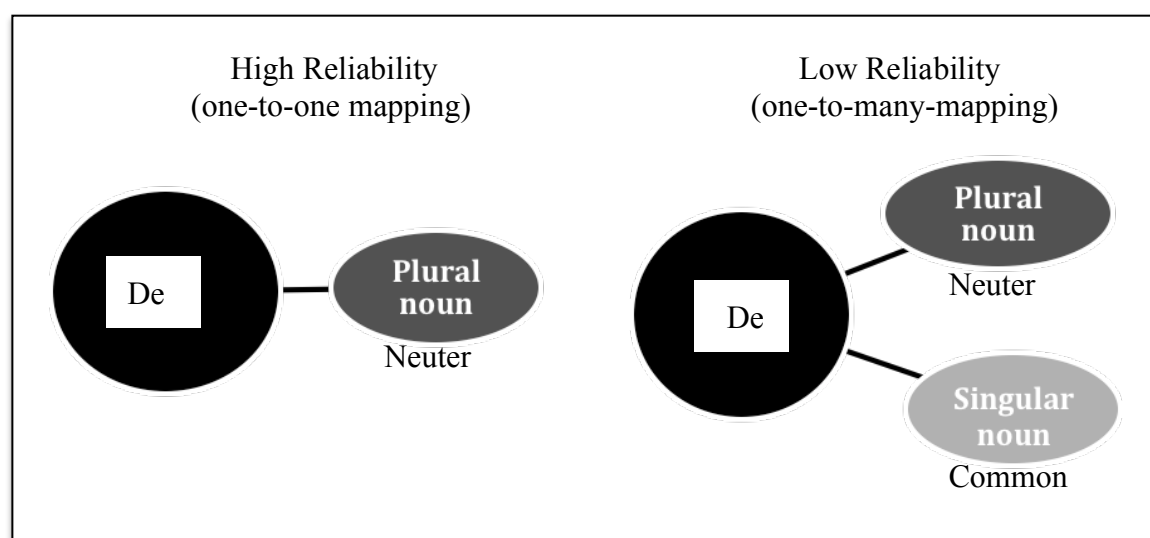


Figure 1. A visual representation of the High and Low reliability mappings of the determiner “de”.

2.4 Participants.

Thirty-two monolingual native English speakers (20 female, 12 male; $M=25.2$ years) with no prior experience with Dutch or German participated in the study. A language history questionnaire was administered to ensure that participants did not speak any other language proficiently. Participants were recruited from the general community in Pittsburgh, Pennsylvania and were compensated \$9 an hour for a total of two hours or received course credit.

2.5 Training Stimuli.

Participants were trained on a set of Dutch stimuli and their English translations over a two-day self-paced training. Stimuli consisted of five determiners, eight nouns—four in the common gender (*de* words) and four in the neuter gender (*het* words), three verbs, one adjective, and one proper name (see Appendix A for stimuli). In addition to being trained on these individual words, participants were also trained on simple Dutch sentences and their English translations, which consisted of various verb-determiner-noun pairings (e.g., *Nicolaas koopt dit schilderij*, ‘Nicholas buys this painting’); these sentences resembled the sentences that participants later saw in their eye-tracking session. Participants saw each word and sentence in its written form while hearing them in their auditory form. All auditory stimuli were recorded by a native Dutch speaker in a sound-attenuated booth. Pictures of the nouns were also presented along with the words and sentences. That is, when hearing and seeing “*dit schilderij* – this painting” or “*Nicolaas koopt dit schilderij* – Nicholas buys this painting,” participants also saw a picture of a painting. All pictures were basic black and white drawings with a black border and were presented at a resolution of 318 x 256 pixels (see Appendix B for example picture stimuli).

Pictures associated with the adjective ‘red’ were colored in with red using Microsoft Paint. The picture stimuli for the Dutch nouns presented during training were identical to those seen during the eye tracking session.

2.6 Training Procedure.

The self-paced training on the computer occurred over two sessions, with one day in between (see Table 2 for study overview); the first session lasted approximately 45 minutes, and the second session lasted one hour and 15 minutes. Vocabulary training occurred in Session 1 and included explicit training on individual verbs and determiners, then the determiners with nouns. All vocabulary items were presented three times, once with pictures and then twice without pictures. Participants saw and heard each Dutch word with its English translation and were asked to repeat the word and its translation. After repeating, participants saw and heard the Dutch lexical item once more for reinforcement, which was followed by a fixation cross to allow participants to move on to the next trial when ready. The presentation of determiners was counterbalanced so that participants learned them in one of three different orders, and the order of noun presentation was randomized for all participants. Given that most determiners trained were of the neuter gender, we included filler training trials of common nouns without their determiners so that there were 24 presentations of each noun in each gender, 12 in the singular form and 12 in the plural form. Number of exposures was balanced so that participants in both reliability conditions were exposed to each determiner and noun the same number of times.

Training Session 2 consisted of a second vocabulary training identical to that of Session 1, plus training on 36 complete sentences (in random order) similar to those they would later hear during the eye-tracking session. Pictures accompanied all vocabulary and sentence items in

Session 2 to reinforce participants' familiarization with them. Immediately before the eye tracking session, participants were asked to complete a verbal translation task in which they saw the written form of a determiner in English (e.g., *this*) next to a pictured noun (e.g., a picture of a painting) and had to translate aloud both the word and the picture into Dutch, for example as *dit schilderij* ("this painting"). Data from participants who were not able to correctly translate a minimum of three of the six determiners were not included in the analyses.

Table 2. *Summary of study procedures*

Session	Procedure	Task
1	Training	Language History Questionnaire Training (Vocabulary with pictures, 1 time; vocabulary without pictures, 2 times)
2	Training	Training (Vocabulary with pictures, 2 times Training (Sentences, 1 time)
	Pre-test Eye tracking	Pre-test (Verbal Translation) Eye tracking test

2.7 Eye-tracking Stimuli.

The eye-tracking stimuli consisted of partial sentences in Dutch (162 sentences in the High reliability condition, 144 in the Low), similar to those presented during training, except that the final noun was removed from the sound files (e.g., *Nicolaas koopt dit _____*, 'Nicholas buys this _____') (cf., Mack et al., 2013). Within each sentence, the determiner constrained the missing noun's gender or number. Twenty-seven partial sentences for each of the six determiners were presented. Each sentence paired the determiner with each verb three times. Each sentence was heard three times, but on each presentation they were paired with one of three different picture pair types in which one, both, or neither picture matched the determiner's gender or number (the

latter two sentence types were treated as fillers and were not included in the analyses). Participants in the Low reliability condition had 18 fewer trials because the determiners *het* and *de* appeared both as *Different* and as *Unique* determiners. For example, for *Different* sentences in which the determiner *de*_{neuterPLUR} is used, participants saw a pair of pictures for which one picture was consistent with that determiner (*de schilderijen* vs. *het raam*), both pictures were consistent (*de schilderijen* vs. *de ramen*) or in which neither was correct (*het schilderij* vs. *het raam*). However, when participants heard *Unique* sentences in which *de*_{commonSING} was used, they would have already seen some of these pairs. To account for this overlap, 18 duplicate sentences were removed, nine for *de* and nine for *het*.

To determine the set of sentences each participant heard and which pairs of pictures appeared with them, four testing lists were created for each reliability condition, and participants were randomly assigned to one of the four lists. Each of the four neuter nouns was paired with each of the other three neuter nouns in a Latin square design; the common nouns were also paired with each other in this same manner. This yielded 12 pairs of pictures, six for which both pictures were of the neuter gender, and six for which both pictures were of the common gender. These 12 within-gender pairs appeared with sentences that required both pictures to be of the same gender. Given that the *Unique* determiners tested a gender distinction, the pictures in these pairs needed to be of opposite gender, so 18 additional opposite-gender pairs were added to both reliability conditions. Although four lists were created, only two contained unique sentences and picture pairs; the other two lists were counterbalanced so that half of the participants would see the target on the left side of the screen, and the other half would see the target on the right.

2.8 Eye-tracking Procedure.

The eye-tracking session took place immediately following the verbal translation task and lasted approximately 20 minutes. Participants were instructed that they would hear partial Dutch sentences while looking at two pictures, and they were asked to indicate with a button press the picture they felt best completed the sentence. A fixation cross appeared at the beginning of each trial, and the participant was instructed to push a button to initial the trial. On each trial the picture pairs appeared 1000ms before sentence onset. The pictures appeared on the left and right of the screen with six inches in between them. Participants' eye movements to the pairs of pictures on the computer were tracked while they listened to the partial Dutch sentences. Eye movements were measured using a desktop version of the Eyelink 1000 eye tracker (SR Research, Mississauga, Ontario, Canada) with a head-stabilized chin rest approximately 22 inches from the camera. Participants in each condition only heard and saw items in one of the four lists, and within this list all items were presented in random order.¹

¹ A second and final block consisted of 24 complete English sentences including a final noun (e.g., 'Click on this book' or 'Click on the coins'); these served as a baseline and were not included in the analysis. A native English speaker recorded these sentences. A different set of black and white pictures of the same resolution accompanied the English sentences.

3.0 RESULTS

3.1 BEHAVIORAL RESULTS.

3.1.1 Accuracy.

A 3 (Similarity) x 2 (Reliability) ANOVA was conducted to compare accuracy in choosing the target picture. Table 3 presents the mean accuracy for each Similarity by Reliability condition. There was a significant main effect of Similarity ($F(2,58) = 40.6, p = .000$), but not of Reliability ($F(1,29) = .61, p = .441$), and there was a significant interaction ($F(2,58) = 8.28, p = .001$).

Table 3. Mean accuracy by Similarity and Reliability condition.

	High	Low
Similar	0.85	0.91
Different	0.78	0.64
Unique	0.44	0.62

To further explore the interaction, we also conducted a Duncan's multiple-range test comparing each Similarity type across Reliability condition (Table 3). This analysis indicated that in the High reliability condition, accuracy for Unique determiners was less than both Similar and Different, whereas in the Low reliability condition, Similar was more accurate than the

Different and Unique determiners. Comparing across reliability conditions, Low Unique was more accurate than High Unique, but Low Different was less accurate than High Different. One-sample t-tests also revealed that each of these Similarity by Reliability conditions was significantly different from chance, with the exception of High Unique ($t(15) = -1.33, p = .21$).

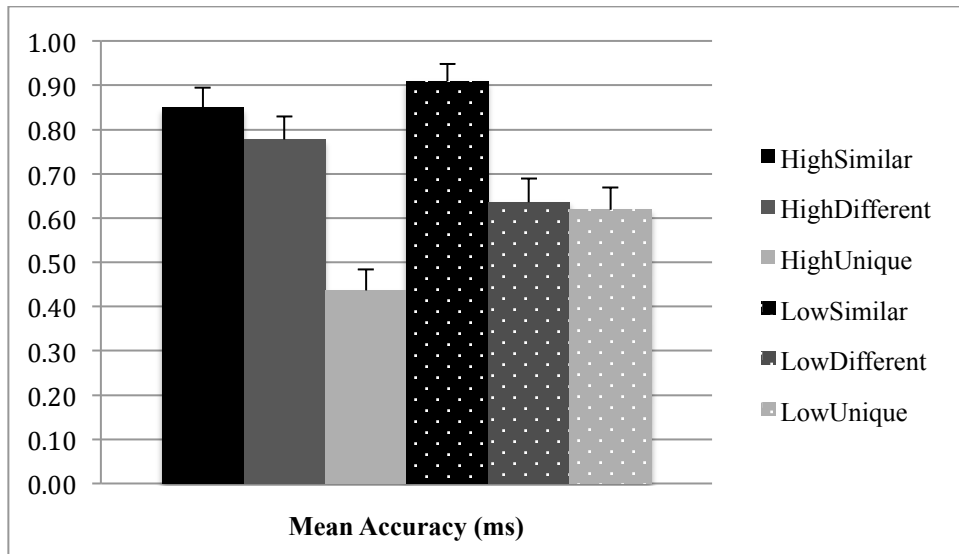


Figure 2. Mean accuracy by Similarity and Reliability condition.

3.1.2 Reaction Time

Mean reaction time (RT) to elicit a button press after determiner offset was also analyzed. Correct trials were included in this analysis (see Table 4 for means). A 3 (Similarity) x 2 (Reliability) ANOVA revealed a significant main effect of Similarity ($F(2,60) = 9.15, p = .000$) and of Reliability ($F(1,30) = 4.56, p = .04$), but no interaction (Figure 3). Follow up t-tests to examine the main effects showed that RTs were faster for Similar sentences compared to Different ($t(31) = -4.4, p = .000$) and for Similar versus Unique sentences ($t(31) = -3.8, p = .001$). RTs in the High reliability group were also faster than for the Low group ($t(15) = -2.29, p = .037$).

Table 4. Mean reaction time (ms) for Similarity and Reliability condition.

	High	Low
Similar	1118.09	1615.16
Different	1656.52	2476.10
Unique	1717.97	2423.21

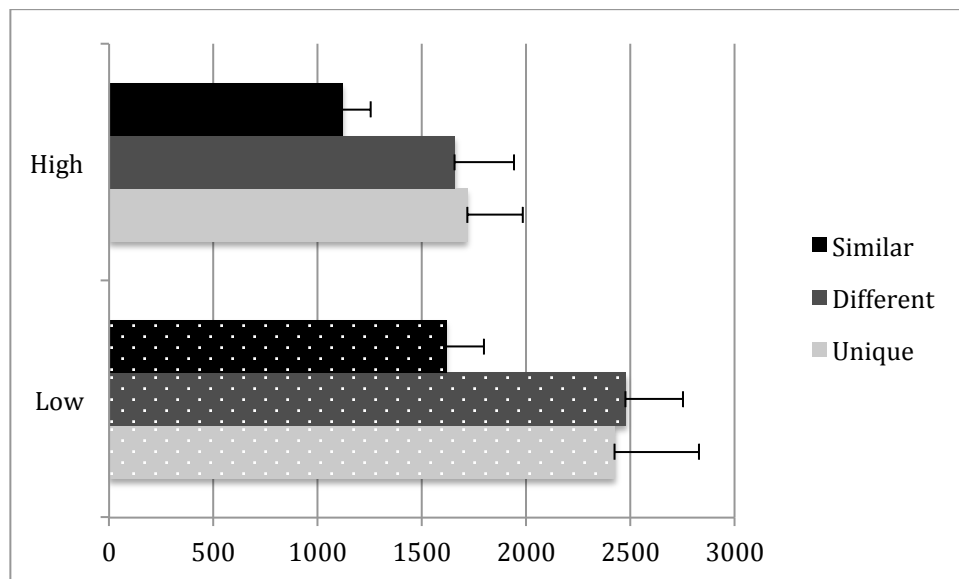


Figure 3. Mean reaction time to elicit a button press after determiner offset by Similarity and Reliability condition.

3.2 EYE-TRACKING RESULTS

3.2.1 First-fixation Latency.

We compared latency of the first fixation to the target picture after determiner offset (correct trials only; Figure 5). Trials in which the first fixation started at or after determiner offset were included; there were no trials on which a first fixation started and ended before determiner offset. A 3 (Similarity) x 2 (Reliability) ANOVA showed a main effect of Similarity

($F(2,58) = 3.15$, $p = .05$), but no significant effect of Reliability ($F(1,29) = 2.11$, $p = .157$) and no interaction ($F(2,58) = .985$, $p = .38$) (Figure 5). Follow-up t-test analyses showed that after determiner offset, participants looked at the target significantly earlier for Similar ($M = 549.5\text{ms}$, $SD = 251.4\text{ms}$) sentences than Unique sentences ($M = 675.7\text{ms}$, $SD = 400.2\text{ms}$; $t(30) = -2.44$, $p = .02$). Latency differences between Similar and Different ($M = 668.2\text{ms}$, $SD = 467.6\text{ms}$) sentences approached significance ($t(31) = -1.914$, $p = .065$), but did not significantly differ in the Different versus Unique sentences (Figure 4).

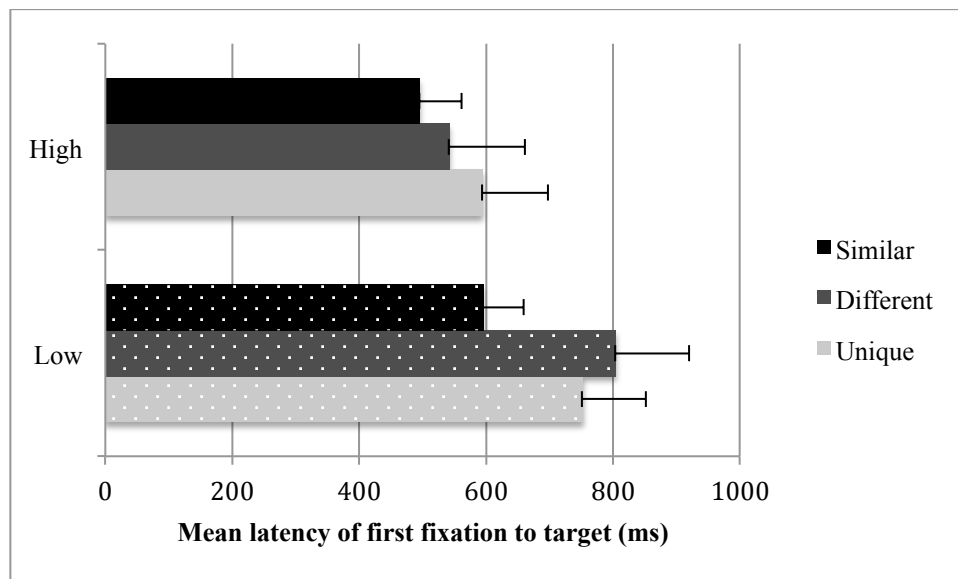


Figure 4. Mean first-fixation latency to look at target from determiner offset for each Reliability and Similarity condition.

3.2.2 Time Course

We examined the time course of fixations to the target picture from 0ms to 2000ms after determiner offset. Figure 5 depicts the cumulative proportion of correct and incorrect trials, divided into 100ms bins, on which participants looked at the target. We excluded trials on which

participants were already looking at the target picture at determiner offset. If a fixation was made to the target within the first 100ms after determiner offset, that trial contributed to the proportion of fixations. Any subsequent fixations to the target in that trial were also included in later time bins, taking into account that a fixation to the target had already occurred in that trial. In efforts to maximize statistical power, we excluded from this analysis five subjects for whom there were fewer than 10 trials in any Similarity condition. For the remaining 27 subjects, cumulative proportion of trials on which participants looked to the target after determiner offset were analyzed using a 3 (Similarity) x 2 (Reliability) x 20 (Time Bin) ANOVA to examine whether there were any differences in changes in the proportion of looking across the entire 2000ms time window. The analysis revealed a main effect of Time ($F(19,475) = 505.48, p = .000$), but no other main effects and no interactions.

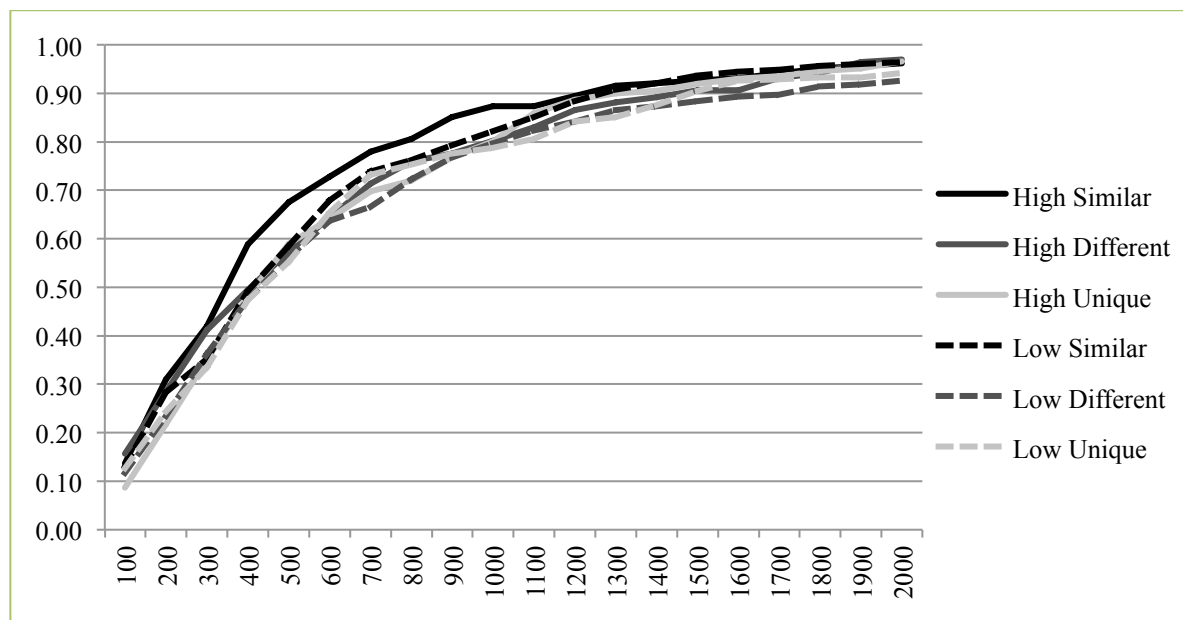


Figure 5. Cumulative proportion of all trials on which participants looked to the target picture after determiner offset for each Similarity and Reliability condition.

We also conducted a second analysis of the cumulative proportion of trials on which participants looked at the target. However, only correct trials were included in this analysis in

order to examine processing that led to correct responses. As we did in the previous analysis, trials in which participants were already looking at the target picture at determiner offset were excluded, as were the same five subjects with fewer than 10 trials in any Similarity condition. This analysis focused on the first 1500ms after determiner offset, after which time all curves appeared to asymptote. Analysis of eye movements within this time window was also important given that most responses were elicited, on average, within 1800ms after determiner offset. The cumulative proportion of trials on which participants looked to the target after determiner offset were analyzed using a 3 (Similarity) x 2 (Reliability) x 15 (Time Bin) ANOVA. We found a significant effect of Time ($F(14,350) = 378.3, p = .000$), and an interaction between Similarity and Time ($F(28,700) = 1.54, p = .038$; Figure 6).

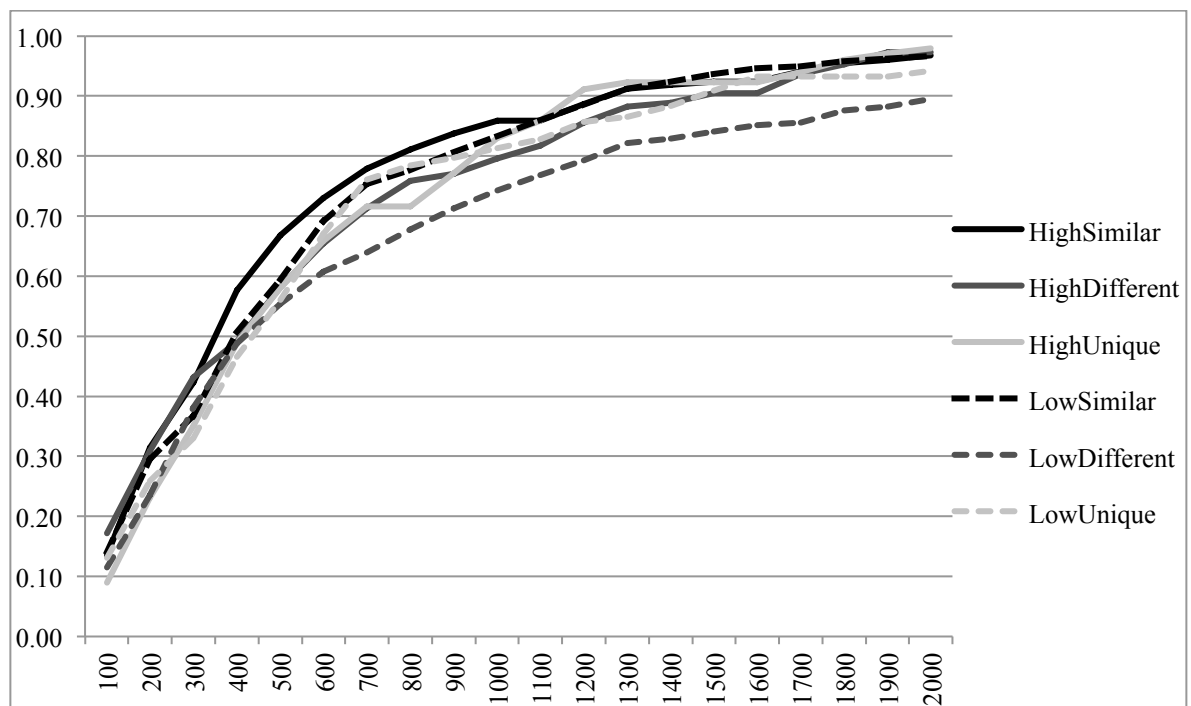


Figure 6. Cumulative proportion of correct trials on which participants looked to the target picture after determiner offset by Similarity and Reliability condition.

Uncorrected t-tests were conducted to examine the earliest point at which there were differences in Similarity conditions. These results revealed that the curve for Similar began to

pull apart from (and is significantly different from) the curve for Unique at the 300-400ms bin and again at the 1300-1400ms bin ($ts(31) \geq 2.07, p < .05$). The curves for Similar and Different pull apart and are significantly different at the 400-500ms bin. These remain significantly different from each other up to the 1400-1500ms bin ($ts(31) \geq -3.09, p < .05$).

4.0 GENERAL DISCUSSION

The current study investigated whether L2 learners are sensitive to morpho-syntactic information on determiners, and whether they use this information online to predict upcoming nouns. Specifically, we examined whether learners' ability to make predictions might be modulated by the similarity in how determiners are instantiated across L1 and L2, by the reliability in how these determiners map to upcoming nouns, or by some interaction between these two variables. To address these questions we examined eye movements to a target picture in partial, determiner-final sentences, and examined RT and accuracy in choosing the target picture.

The framework of the UCM postulates that linguistic structures in the L2 that work similarly in the L1 benefit from transfer and are easiest to learn. Structures that exist in both languages, but that are implemented differently in both languages give rise to competition and are hardest to learn. Determiners unique to the L2 do not give rise to transfer or competition, so the learnability for Unique determiners should be determined by their reliability as well as availability in the input.

The present study is unique because it examines not only the individual effects of cross-language similarity and reliability, but also how the effect of these two mechanisms might interact. In early L2 acquisition, comprehenders are likely to encounter linguistic information that are characterized by varying levels of similarity and reliability mappings, so it is useful to

understand how these morpho-syntactic features might affect their learning as well as how they use this information during comprehension.

The behavioral results for accuracy support the predictions of the UCM. We found a main effect of Similarity as well as a significant interaction between Similarity and Reliability. We expected that accuracy would be higher for the Different and Unique determiners in the High reliability condition than Different and Unique determiners in the Low condition. Surprisingly, pairwise comparisons for accuracy showed that accuracy for Unique determiners was higher in the Low reliability condition than in the High condition. This suggests that the *rood/rode*-to-noun mapping (gender distinction, High condition) was more difficult to learn than the *het/de*_{commonSING}-to-noun mapping (gender distinction, Low condition). It is possible that there were other concerns other than reliability that may have affected the learnability of the Unique determiners. For instance, regardless of the duplicate mapping of *de*, if a participant in the Low condition had at the very least learned the *het*-to-noun mapping (a highly reliable mapping in our study), it could have led them to get the sentences with Unique determiners correct about half the time. However, if participants in the High condition's sensitivity to morpho-syntactic cues such as the long vs. short vowel distinction and/or the *-e* inflection in *rood/rode* was weak, then not only would this distinction and its noun mapping have been harder to learn, the chances of correctly choosing the target in sentences would have been lowered.

Other pairwise comparisons for accuracy were more in line with our hypotheses. For example, performance on the Different determiners was more accurate in the High condition than in the Low condition. The Different determiners were identical in both reliability conditions, so this is evidence that in the Low group, the learnability of the Different determiners was affected by the low reliability of the mapping of the determiner *de* as being both a Different (*de*_{neuterPLUR})

and Unique determiner ($de_{\text{commonSING}}$). As for within group comparisons, in the High reliability condition Unique was less accurate than both Similar and Different determiners and in the Low reliability condition, Unique was less accurate than Similar determiners. These results support the predictions of the UCM in that L1-L2 similarity in determiner instantiation modulated accuracy in choosing the target picture. Participants most accurately chose the target picture when the determiner was instantiated similarly across L1 and L2 (*dit* and *deze*), suggesting that learnability of these types of determiners benefited from transfer from the L1.

Cross-language similarity also modulated how quickly participants selected a picture as their response. A main effect of Similarity revealed that RTs for Similar determiners were faster than for Unique determiners, and only marginally faster than RTs for Different determiners, consistent with the UCM. Interestingly, although RTs for Different determiners were as slow as for Unique determiners, accuracy was higher overall for Different determiners than Unique determiners. This might be evidence that learners in both conditions encountered L1-L2 competition in processing the Different determiners. That is, although participants were slower in their use of morpho-syntactic cues on the Different determiners, they still more accurately chose the target in sentences with Different determiners than Unique determiners. So although more time was needed to process the information on the Different determiners, it paid off in higher accuracy. RT data also showed a main effect of reliability such that RTs for the High reliability condition were faster than for the Low reliability condition. This supports the UCM's predictions for the effects of reliability in determiner-noun mappings. Of note, there is an overall mismatch between RT and accuracy data when looking at the Reliability main effect, such that significantly faster RTs in the High vs. Low condition did not result in any accuracy differences between reliability groups.

Overall, the behavioral data support some of the predictions of the UCM. Accuracy and RT data suggest that the learnability of determiner-noun mappings was modulated by cross-language similarity. It also suggests that participants more efficiently used morpho-syntactic cues on the determiner when they are instantiated similarly across both languages.

Eye tracking results were not as consistent with the predictions of the UCM as were the behavioral data. We examined the latencies of the first fixation to the target picture after determiner offset and found a main effect of similarity, but no effects of reliability and no interaction between the two variables. Participants oriented to the target picture faster when the determiners were instantiated similarly across English and Dutch than when the determiners were unique to Dutch. Latency of first fixation for Different determiners fell somewhere in between these two such that their latencies were not significantly different from the latencies for the Unique determiners, but approached significance in their difference from the Similar determiners. This suggests that participants used morpho-syntactic information on the determiners to orient to the target picture after hearing the determiner, and did so most efficiently for determiners that are instantiated similarly across L1 and L2. This sensitivity to morpho-syntactic information was a key finding and is evidence that cross-language similarity influenced the success of learning the different determiner-noun mappings. We expected that participants would be faster to orient to the target for Different and Unique determiners in the High reliability condition than for Different and Unique determiners in the Low reliability condition. However, we found no evidence that cross-language similarity and reliability interacted to influence on-line processing.

We also evaluated the time course of fixations, examining the proportion of correct trials (within 0-1500ms) on which there was a fixation to the target. Analyses focused within this time

frame because on average, participants elicited a button press around 1800ms for all three determiner types. We found a significant interaction of Similarity and Time. Visually, this pattern is evident, as we can see the curves for Similar, Different and Unique start to pull apart starting as early as within 300-400ms after determiner offset, which might be evidence that learners are immediately using the information even though they did not respond until much later. We can interpret the eye tracking data as suggesting that the information on the determiner was a helpful cue to learners in orienting to the target picture, and was most helpful when the determiner is instantiated similarly across L1-L2. Examining accurate trials only allows us to focus on processing that led to correct responses, which will essentially provide insight into what information is useful in getting the correct answer rather than what information is used more generally. However, these results must be interpreted with caution given that accuracy for High Unique was not significantly different from chance. On the other hand, the analysis of all trials might have given us a better picture of participants' overall implicit processing that is unfiltered by explicit performance. No significant effects were found when the proportion of fixations were considered for all trials, which suggests that any effects might have been washed out when diffused over all trials. Furthermore, the significant patterns in the proportional eye movement data (when considering correct trials only) are consistent with eye movement latency patterns as well as with the behavioral data. This consistency only corroborates the results and provides a coherent picture of how learners are using the information on the determiner. It also demonstrates that learners have access to both implicit and explicit knowledge online, and that both contribute to making predictions.

Overall, results are consistent with the UCM such that robust effects of cross-language similarity were evident in all analyses. Reliability effects, however, were not as consistent in the

data, but the effects found were interesting. It is possible that our reliability outcomes were affected by the apparent difficulty of learning the *rood/rode* determiners compared to the *het/de* distinction in the Low Unique condition. The UCM posits that learning determiners unique to the L2 relies on its strength as a cue, which includes its reliability, but also its availability. In addition to fact that inflection is not a strong cue in English, this type of morpho-syntactic feature (long vs. short vowel accompanied by an *-e* inflection) is not characteristic of the Romance Languages that the typical native English speaker in our sample might have been previously exposed to. The experimenters were careful to make the long and short vowel differences and the *-e* inflection as clear as possible during the recording of the stimuli. However, it seems very likely that this unique feature was not attended to or went unnoticed (Ellis, 2006) given that our training, although providing explicit training, was mostly exposure-based. Tolentino and Tokowicz (2014) suggest that perhaps increased salience or explicitly highlighting morpho-syntactic rules in training might have helped learners attend to and better learn the morpho-syntactic information on Unique determiners. This might have increased cue strength and learning of the *rood/rode*-to-noun mapping, and perhaps stronger effects of Reliability might have emerged, particularly ones that are consistent with the UCM.

In conclusion, our results indicate that learners were sensitive to morpho-syntactic information on determiners and were able to make predictions of upcoming nouns on-line using these determiners. Our covert-noun paradigm allowed us to examine predictive eye movements within a time frame that in previous studies might have coincided with the presentation of a noun. By giving learners more time to make a prediction we were also able to capture a clearer picture of what information learners make use of during comprehension. Examination of these processes supports the idea that learners used both explicit and implicit processes on-line and

that these both contribute to making predictions. Future work should explore the nature of the relationship between these explicit and implicit processes and what it might mean for predictions if evidence is present for one without the other. In sum, the results suggests that cross-language similarity modulates the learnability of mapping a determiner to its noun, and that beginning L2 learners can use this morpho-syntactic information to make predictions during on-line sentence comprehension.

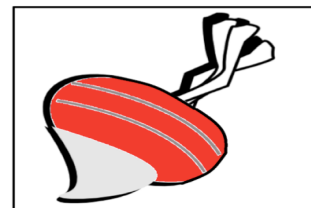
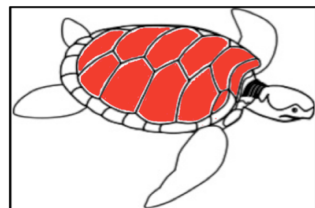
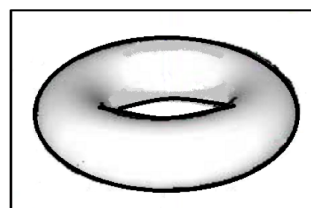
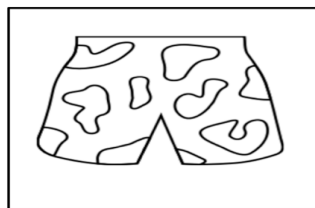
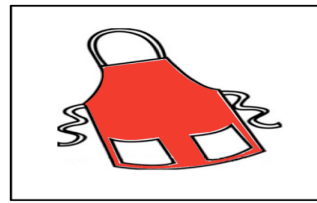
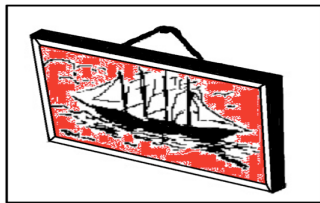
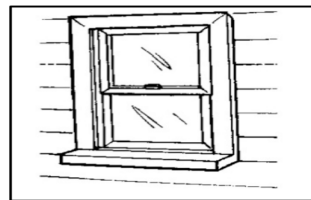
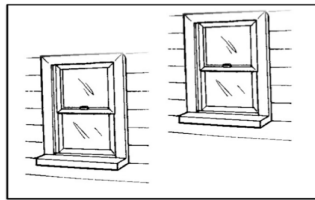
APPENDIX A

Table 5. *List of word stimuli.*

Determiners	Nouns	Verb	Proper name
<u>Similar</u>	<u>Neuter</u>	<i>raakt</i> (touches)	Nicolaas
Dit (this)	Het schilderij (the painting)	<i>koopt</i> (buys)	
Deze (these)	Het zwempak (the swimsuit)	<i>ziet</i> (sees)	
Het keukenschort (the apron)			
Het raam (the window)			
<u>Different</u>	<u>Common</u>		
Het _{neuterSING} (the _{SING})	De schildpad (the turtle)		
De _{neuterPLUR} (the _{PLUR})	De zwemband (the inner tube)		
	De keukentrap (the stepladder)		
	De raap (the turnip)		
<u>Unique (High)</u>			
Een rode _{common} /rood _{neuter} (a red)			
<u>Unique (Low)</u>			
De _{commonSING} (the _{SING})			
Het _{neuterSING} (the _{SING})			

APPENDIX B

EXAMPLES OF SIDE-BY-SIDE PICTURE PAIRS SEEN DURING THE EYE-TRACKING TEST.



BIBLIOGRAPHY

- Altmann, G.T.M., & Kamide, Y. (1999). Incremental interpretation at verbs: restricting the domain of subsequent reference. *Cognition*, 73(3), 247-264.
- Bates, E., Devescovi, A., Hernandez, A., & Pizzamiglio, L. (1996). Gender priming in Italian. *Perception & Psychophysics*, 85(7), 992-1004.
- Bates, E. & MacWhinney, B. (1987). Competition, variation, and language learning. In B. MacWhinney (Ed.). *Mechanisms of language acquisition*. London, UK: Lawrence Erlbaum Associates, Inc.
- Borovsky, A., Elmann, J.L. & Fernald, A. (2012). Knowing a lot for one's age: Vocabulary skill and not age is associated with anticipatory incremental sentence interpretation in children and adults. *Journal of Experimental Child Psychology*, 112(4), 417-436.
- Brunellière, A. & Soto-Faraco, S. (2013). The speakers' accent shapes the listeners' phonological predictions during speech perception. *Brain and Language*, 125, 82-93.
- Dahan, D., Swingle, D., Tanenhaus, M.K. & Magnuson, J.S. (2000). Linguistic gender and spoken-word recognition in French. *Journal of Memory and Language*, 42, 465-480.
- Dussias, P. E. (2010). Uses of eyetracking data in second language sentence processing research. *Annual Review of Applied Linguistics*, 30, 149 -166.
- Foucart, A. & Frenck-Mestre, C. (2011). Grammatical gender processing in L2: electrophysiological evidence of the effect of the L1-L2 syntactic similarity. *Bilingualism: Language and Cognition*, 14(3), 379-399.
- Gillon Dowens, M., Guo, T., Guo, J., Barber, H. & Carreiras, M. (2011). Gender and number processing in Chinese learners of Spanish – Evidence from Event Related Potentials. *Neuropsychologia*, 49, 1651-1659
- Grosjean F., Dommergues, J.Y., Cornu, E., Guillelmon, D. & Besson, C. (1994). The gender marking effect in spoken word recognition. *Perception & Psychophysics*, 56, 590–598.
- Grüter, T., Lew-Williams, C. & Fernald, A. (2011). Grammatical Gender in L2: A production or real-time processing problem? *Second Language Research*, 28(2), 191-215.

- Kamide, Y., Altmann, G.T.M., & Haywood, S.L. (2003). The time-course of prediction in incremental sentence processing: Evidence from anticipatory eye movements. *Journal of Memory and Language*, 49(1), 133-156.
- Lew-Williams, C. & Fernald, A. (2007). Young children learning Spanish make rapid use of grammatical gender in spoken word recognition. *Psychological Science*, 18, 193-198.
- Lew-Williams, C. & Fernald, A. (2010). Real-time processing of gender-marked articles by native and non-native Spanish speakers. *Journal of Memory and Language*, 63(4), 447-464.
- Mack, J., Woohyuk J., Thompson, C. (2013). Effects of verb meaning on lexical integration in agrammatic aphasia: Evidence from eye tracking. *Journal of Neurolinguistics* 26, 619-636.
- MacWhinney, B. (1987). Applying the Competition Model to bilingualism. *Applied Psycholinguistics*, 8, 315-327.
- MacWhinney, B. (2005). Extending the Competition Model. *International Journal of Bilingualism*, 9(7), 69-84.
- McDonald, J. L. (1987). Sentence interpretation in bilingual speakers of English and Dutch. *Applied Psycholinguistics*, 8, 379-413.
- Sabourin, L., Stowe, L.A. & de Haan, Ger J. (2006). Transfer effects in learning a second language grammatical gender system. *Second Language Research*, 22(1), 1-29.
- Tokowicz, N. & Warren, T. (2010). Beginning adult L2 learner's sensitivity to morphosyntactic violations: A self-paced reading study. *European Journal of Cognitive Psychology*, 22(2), 1092-1106.
- Tokowicz, N. & MacWhinney, B. (2005). Implicit and explicit measures of sensitivity to violations in second language grammar. *Studies in Second Language Acquisition*, 27, 173-204.
- Tolentinto, L. & Tokowicz, N. (2014). Cross-language similarity modulates effectiveness of second language grammar instruction. *Language Learning*, 64(2), 279-309.
- Valdés Kroff, J., Guzzardo, R., Dussias, P. E., Gerfen, C., Gullifer, J. (2008). *Processing spoken code-switches by Spanish-English bilinguals: A visual world study*. Paper presented at the 7th International Symposium on Bilingualism, Utrecht, The Netherlands.
- Van Berkum, J.J.A, Brown, C.M., Zwitserlood, P., Kooijman, V. & Hagoort, P. (2005). Anticipating Upcoming Words in Discourse: Evidence from ERPs and Reading Times. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 31(3), 443-467.

Wicha, N. Y. Y., Moreno, E. M., & Kutas, M. (2004). Anticipating words and their gender: An event-related brain potential study of semantic integration, gender expectancy, and gender agreement in Spanish sentence reading. *Journal of Cognitive Neuroscience* 16(7), 1272-1288.